

(19)



Europäisches Patentamt

European Patent Office

Office européen des brevets



(11)

EP 0 879 881 A1

(12)

EUROPEAN PATENT APPLICATION

(43) Date of publication:
25.11.1998 Bulletin 1998/48

(51) Int. Cl.⁶: **C12N 15/12**, **C07K 14/71**,
C12P 21/00

(21) Application number: **97308556.6**

(22) Date of filing: **27.10.1997**

(84) Designated Contracting States:
**AT BE CH DE DK ES FI FR GB GR IE IT LI LU MC
NL PT SE**

(30) Priority: **23.05.1997 US 47626 P**
14.07.1997 US 892083

(71) Applicant:
SMITHKLINE BEECHAM CORPORATION
Philadelphia Pennsylvania 19103 (US)

(72) Inventors:
• **Hu, Erding,**
SmithKline Beecham Pharma.
King of Prussia, Pennsylvania 19406 (US)
• **Zhu, Yuan,**
SmithKline Beecham Pharma.
King of Prussia, Pennsylvania 19406 (US)

(74) Representative:
Crump, Julian Richard John et al
FJ Cleveland,
40-43 Chancery Lane
London WC2A 1JQ (GB)

(54) **A human gene similar to a secreted protein frizb (ATG-1639)**

(57) ATG-1639 polypeptides and polynucleotides and methods for producing such polypeptides by recombinant techniques are disclosed. Also disclosed are methods for utilizing ATG-1639 polypeptides and polynucleotides in the design of protocols for the treatment of heart disease, hypertension, cardiovascular diseases, kidney diseases, obesity, insulin resistance, diabetes, CNS diseases, additions., among others, and diagnostic assays for such conditions.

EP 0 879 881 A1

Description**FIELD OF INVENTION**

5 This invention relates to newly identified polynucleotides, polypeptides encoded by them and to the use of such polynucleotides and polypeptides, and to their production. More particularly, the polynucleotides and polypeptides of the present invention relate to secreted protein ligand for 7-TM receptor, frizzled family, hereinafter referred to as ATG-1639. The invention also relates to inhibiting or activating the action of such polynucleotides and polypeptides.

10 BACKGROUND OF THE INVENTION

Protein ligands for 7-TM receptors are important molecules that regulate a variety of biological effects. Glucagon, morphin, opiate, and thrombin are all examples of these protein ligands. Glucagon, for example, elevates blood glucose levels by interacting with the glucagon receptor, a 7-TM receptor, and transducing signals to the cellular glucose regu-
 15 latory machinery. Thus, identification of a novel protein ligand for 7-TM receptors may reveal novel pathways to many biological processes in normal and pathological states. This indicates that the secreted protein ligand for 7-TM receptor, frizzled family has an established, proven history as therapeutic targets. Clearly there is a need for identification and characterization of further members of secreted protein ligand for 7-TM receptor, frizzled family which can play a role in preventing, ameliorating or correcting dysfunctions or diseases, including, but not limited to, heart disease, hyperten-
 20 sion, cardiovascular diseases, kidney diseases, obesity, insulin resistance, diabetes, CNS diseases, additions..

SUMMARY OF THE INVENTION

In one aspect, the invention relates to ATG-1639 polypeptides and recombinant materials and methods for their
 25 production. Another aspect of the invention relates to methods for using such ATG-1639 polypeptides and polynucleotides. Such uses include the treatment of heart disease, hypertension, cardiovascular diseases, kidney diseases, obesity, insulin resistance, diabetes, CNS diseases, additions., among others. In still another aspect, the invention relates to methods to identify agonists and antagonists using the materials provided by the invention, and treating conditions associated with ATG-1639 imbalance with the identified compounds. Yet another aspect of the invention relates
 30 to diagnostic assays for detecting diseases associated with inappropriate ATG-1639 activity or levels.

DESCRIPTION OF THE INVENTION**Definitions**

35 The following definitions are provided to facilitate understanding of certain terms used frequently herein.

"ATG-1639" refers, among others, generally to a polypeptide having the amino acid sequence set forth in SEQ ID NO:2 or an allelic variant thereof.

40 "ATG-1639 activity or ATG-1639 polypeptide activity" or "biological activity of the ATG-1639 or ATG-1639 polypeptide" refers to the metabolic or physiologic function of said ATG-1639 including similar activities or improved activities or these activities with decreased undesirable side-effects. Also included are antigenic and immunogenic activities of said ATG-1639.

"ATG-1639 gene" refers to a polynucleotide having the nucleotide sequence set forth in SEQ ID NO:1 or allelic variants thereof and/or their complements.

45 "Antibodies" as used herein includes polyclonal and monoclonal antibodies, chimeric, single chain, and humanized antibodies, as well as Fab fragments, including the products of an Fab or other immunoglobulin expression library.

"Isolated" means altered "by the hand of man" from the natural state. If an "isolated" composition or substance occurs in nature, it has been changed or removed from its original environment, or both. For example, a polynucleotide or a polypeptide naturally present in a living animal is not "isolated," but the same polynucleotide or polypeptide separated from the coexisting materials of its natural state is "isolated", as the term is employed herein.
 50

"Polynucleotide" generally refers to any polyribonucleotide or polydeoxribonucleotide, which may be unmodified RNA or DNA or modified RNA or DNA. "Polynucleotides" include, without limitation single- and double-stranded DNA, DNA that is a mixture of single- and double-stranded regions, single- and double-stranded RNA, and RNA that is mixture of single- and double-stranded regions, hybrid molecules comprising DNA and RNA that may be single-stranded or,
 55 more typically, double-stranded or a mixture of single- and double-stranded regions. In addition, "polynucleotide" refers to triple-stranded regions comprising RNA or DNA or both RNA and DNA. The term polynucleotide also includes DNAs or RNAs containing one or more modified bases and DNAs or RNAs with backbones modified for stability or for other reasons. "Modified" bases include, for example, tritylated bases and unusual bases such as inosine. A variety of modi-

fications has been made to DNA and RNA; thus, "polynucleotide" embraces chemically, enzymatically or metabolically modified forms of polynucleotides as typically found in nature, as well as the chemical forms of DNA and RNA characteristic of viruses and cells. "Polynucleotide" also embraces relatively short polynucleotides, often referred to as oligonucleotides.

"Polypeptide" refers to any peptide or protein comprising two or more amino acids joined to each other by peptide bonds or modified peptide bonds, i.e., peptide isosteres. "Polypeptide" refers to both short chains, commonly referred to as peptides, oligopeptides or oligomers, and to longer chains, generally referred to as proteins. Polypeptides may contain amino acids other than the 20 gene-encoded amino acids. "Polypeptides" include amino acid sequences modified either by natural processes, such as posttranslational processing, or by chemical modification techniques which are well known in the art. Such modifications are well described in basic texts and in more detailed monographs, as well as in a voluminous research literature. Modifications can occur anywhere in a polypeptide, including the peptide backbone, the amino acid side-chains and the amino or carboxyl termini. It will be appreciated that the same type of modification may be present in the same or varying degrees at several sites in a given polypeptide. Also, a given polypeptide may contain many types of modifications. Polypeptides may be branched as a result of ubiquitination, and they may be cyclic, with or without branching. Cyclic, branched and branched cyclic polypeptides may result from posttranslation natural processes or may be made by synthetic methods. Modifications include acetylation, acylation, ADP-ribosylation, amidation, covalent attachment of flavin, covalent attachment of a heme moiety, covalent attachment of a nucleotide or nucleotide derivative, covalent attachment of a lipid or lipid derivative, covalent attachment of phosphatidylinositol, cross-linking, cyclization, disulfide bond formation, demethylation, formation of covalent cross-links, formation of cystine, formation of pyroglutamate, formylation, gamma-carboxylation, glycosylation, GPI anchor formation, hydroxylation, iodination, methylation, myristoylation, oxidation, proteolytic processing, phosphorylation, prenylation, racemization, selenoylation, sulfation, transfer-RNA mediated addition of amino acids to proteins such as arginylation, and ubiquitination. See, for instance, PROTEINS - STRUCTURE AND MOLECULAR PROPERTIES, 2nd Ed., T. E. Creighton, W. H. Freeman and Company, New York, 1993 and Wold, F., Posttranslational Protein Modifications: Perspectives and Prospects, pgs. 1-12 in POSTTRANSLATIONAL COVALENT MODIFICATION OF PROTEINS, B. C. Johnson, Ed., Academic Press, New York, 1983; Seifter *et al.*, "Analysis for protein modifications and nonprotein cofactors", *Meth Enzymol* (1990) 182:626-646 and Rattan *et al.*, "Protein Synthesis: Posttranslational Modifications and Aging", *Ann NY Acad Sci* (1992) 663:48-62.

"Variant" as the term is used herein, is a polynucleotide or polypeptide that differs from a reference polynucleotide or polypeptide respectively, but retains essential properties. A typical variant of a polynucleotide differs in nucleotide sequence from another, reference polynucleotide. Changes in the nucleotide sequence of the variant may or may not alter the amino acid sequence of a polypeptide encoded by the reference polynucleotide. Nucleotide changes may result in amino acid substitutions, additions, deletions, fusions and truncations in the polypeptide encoded by the reference sequence, as discussed below. A typical variant of a polypeptide differs in amino acid sequence from another, reference polypeptide. Generally, differences are limited so that the sequences of the reference polypeptide and the variant are closely similar overall and, in many regions, identical. A variant and reference polypeptide may differ in amino acid sequence by one or more substitutions, additions, deletions in any combination. A substituted or inserted amino acid residue may or may not be one encoded by the genetic code. A variant of a polynucleotide or polypeptide may be a naturally occurring such as an allelic variant, or it may be a variant that is not known to occur naturally. Non-naturally occurring variants of polynucleotides and polypeptides may be made by mutagenesis techniques or by direct synthesis.

"Identity" is a measure of the identity of nucleotide sequences or amino acid sequences. In general, the sequences are aligned so that the highest order match is obtained. "Identity" *per se* has an art-recognized meaning and can be calculated using published techniques. See, e.g.: (COMPUTATIONAL MOLECULAR BIOLOGY, Lesk, A.M., ed., Oxford University Press, New York, 1988; BIOCOMPUTING: INFORMATICS AND GENOME PROJECTS, Smith, D.W., ed., Academic Press, New York, 1993; COMPUTER ANALYSIS OF SEQUENCE DATA, PART I, Griffin, A.M., and Griffin, H.G., eds., Humana Press, New Jersey, 1994; SEQUENCE ANALYSIS IN MOLECULAR BIOLOGY, von Heinje, G., Academic Press, 1987; and SEQUENCE ANALYSIS PRIMER, Gribskov, M. and Devereux, J., eds., M Stockton Press, New York, 1991). While there exist a number of methods to measure identity between two polynucleotide or polypeptide sequences, the term "identity" is well known to skilled artisans (Carillo, H., and Lipton, D., *SIAM J Applied Math* (1988) 48:1073). Methods commonly employed to determine identity or similarity between two sequences include, but are not limited to, those disclosed in Guide to Huge Computers, Martin J. Bishop, ed., Academic Press, San Diego, 1994, and Carillo, H., and Lipton, D., *SIAM J Applied Math* (1988) 48:1073. Methods to determine identity and similarity are codified in computer programs. Preferred computer program methods to determine identity and similarity between two sequences include, but are not limited to, GCS program package (Devereux, J., *et al.*, *Nucleic Acids Research* (1984) 12(1):387), BLASTP, BLASTN, FASTA (Atschul, S.F. *et al.*, *J Molec Biol* (1990) 215:403).

As an illustration, by a polynucleotide having a nucleotide sequence having at least, for example, 95% "identity" to a reference nucleotide sequence of SEQ ID NO: 1 is intended that the nucleotide sequence of the polynucleotide is

identical to the reference sequence except that the polynucleotide sequence may include up to five point mutations per each 100 nucleotides of the reference nucleotide sequence of SEQ ID NO: 1. In other words, to obtain a polynucleotide having a nucleotide sequence at least 95% identical to a reference nucleotide sequence, up to 5% of the nucleotides in the reference sequence may be deleted or substituted with another nucleotide, or a number of nucleotides up to 5% of the total nucleotides in the reference sequence may be inserted into the reference sequence. These mutations of the reference sequence may occur at the 5 or 3 terminal positions of the reference nucleotide sequence or anywhere between those terminal positions, interspersed either individually among nucleotides in the reference sequence or in one or more contiguous groups within the reference sequence.

Similarly, by a polypeptide having an amino acid sequence having at least, for example, 95% "identity" to a reference amino acid sequence of SEQ ID NO:2 is intended that the amino acid sequence of the polypeptide is identical to the reference sequence except that the polypeptide sequence may include up to five amino acid alterations per each 100 amino acids of the reference amino acid of SEQ ID NO: 2. In other words, to obtain a polypeptide having an amino acid sequence at least 95% identical to a reference amino acid sequence, up to 5% of the amino acid residues in the reference sequence may be deleted or substituted with another amino acid, or a number of amino acids up to 5% of the total amino acid residues in the reference sequence may be inserted into the reference sequence. These alterations of the reference sequence may occur at the amino or carboxy terminal positions of the reference amino acid sequence or anywhere between those terminal positions, interspersed either individually among residues in the reference sequence or in one or more contiguous groups within the reference sequence.

Polypeptides of the Invention

In one aspect, the present invention relates to ATG-1639 polypeptides. The ATG-1639 polypeptides include the polypeptide of SEQ ID NOS:2 and 4; as well as polypeptides comprising the amino acid sequence of SEQ ID NO: 2; and polypeptides comprising the amino acid sequence which have at least 80% identity to that of SEQ ID NO:2 over its entire length, and still more preferably at least 90% identity, and even still more preferably at least 95% identity to SEQ ID NO: 2. Furthermore, those with at least 97-99% are highly preferred. Also included within ATG-1639 polypeptides are polypeptides having the amino acid sequence which have at least 80% identity to the polypeptide having the amino acid sequence of SEQ ID NO:2 over its entire length, and still more preferably at least 90% identity, and still more preferably at least 95% identity to SEQ ID NO:2. Furthermore, those with at least 97-99% are highly preferred. Preferably ATG-1639 polypeptide exhibit at least one biological activity of ATG-1639.

The ATG-1639 polypeptides may be in the form of the "mature" protein or may be a part of a larger protein such as a fusion protein. It is often advantageous to include an additional amino acid sequence which contains secretory or leader sequences, pro-sequences, sequences which aid in purification such as multiple histidine residues, or an additional sequence for stability during recombinant production.

Fragments of the ATG-1639 polypeptides are also included in the invention. A fragment is a polypeptide having an amino acid sequence that entirely is the same as part, but not all, of the amino acid sequence of the aforementioned ATG-1639 polypeptides. As with ATG-1639 polypeptides, fragments may be "free-standing," or comprised within a larger polypeptide of which they form a part or region, most preferably as a single continuous region. Representative examples of polypeptide fragments of the invention, include, for example, fragments from about amino acid number 1-20, 21-40, 41-60, 61-80, 81-100, and 101 to the end of ATG-1639 polypeptide. In this context "about" includes the particularly recited ranges larger or smaller by several, 5, 4, 3, 2 or 1 amino acid at either extreme or at both extremes.

Preferred fragments include, for example, truncation polypeptides having the amino acid sequence of ATG-1639 polypeptides, except for deletion of a continuous series of residues that includes the amino terminus, or a continuous series of residues that includes the carboxyl terminus or deletion of two continuous series of residues, one including the amino terminus and one including the carboxyl terminus. Also preferred are fragments characterized by structural or functional attributes such as fragments that comprise alpha-helix and alpha-helix forming regions, beta-sheet and beta-sheet-forming regions, turn and turn-forming regions, coil and coil-forming regions, hydrophilic regions, hydrophobic regions, alpha amphipathic regions, beta amphipathic regions, flexible regions, surface-forming regions, substrate binding region, and high antigenic index regions. Other preferred fragments are biologically active fragments. Biologically active fragments are those that mediate ATG-1639 activity, including those with a similar activity or an improved activity, or with a decreased undesirable activity. Also included are those that are antigenic or immunogenic in an animal, especially in a human.

Preferably, all of these polypeptide fragments retain the biological activity of the ATG-1639, including antigenic activity. Among the most preferred fragment is that having the amino acid sequence of SEQ ID NO: 4. Variants of the defined sequence and fragments also form part of the present invention. Preferred variants are those that vary from the referents by conservative amino acid substitutions -- i.e., those that substitute a residue with another of like characteristics. Typical such substitutions are among Ala, Val, Leu and Ile; among Ser and Thr; among the acidic residues Asp and Glu; among Asn and Gln; and among the basic residues Lys and Arg; or aromatic residues Phe and Tyr. Particularly

preferred are variants in which several, 5-10, 1-5, or 1-2 amino acids are substituted, deleted, or added in any combination.

The ATG-1639 polypeptides of the invention can be prepared in any suitable manner. Such polypeptides include isolated naturally occurring polypeptides, recombinantly produced polypeptides, synthetically produced polypeptides, or polypeptides produced by a combination of these methods. Means for preparing such polypeptides are well understood in the art.

Polynucleotides of the Invention

Another aspect of the invention relates to ATG-1639 polynucleotides. ATG-1639 polynucleotides include isolated polynucleotides which encode the ATG-1639 polypeptides and fragments, and polynucleotides closely related thereto. More specifically, ATG-1639 polynucleotide of the invention include a polynucleotide comprising the nucleotide sequence set forth in SEQ ID NO:1 encoding a ATG-1639 polypeptide of SEQ ID NO: 2, and polynucleotides having the particular sequences of SEQ ID NOS:1 and 3. ATG-1639 polynucleotides further include a polynucleotide comprising a nucleotide sequence that has at least 80% identity to a nucleotide sequence encoding the ATG-1639 polypeptide of SEQ ID NO:2 over its entire length, and a polynucleotide that is at least 80% identical to that having SEQ ID NO:1 over its entire length. In this regard, polynucleotides at least 90% identical are particularly preferred, and those with at least 95% are especially preferred. Furthermore, those with at least 97% are highly preferred and those with at least 98-99% are most highly preferred, with at least 99% being the most preferred. Also included under ATG-1639 polynucleotides are a nucleotide sequence which has sufficient identity to a nucleotide sequence contained in SEQ ID NO:1 to hybridize under conditions useable for amplification or for use as a probe or marker. The invention also provides polynucleotides which are complementary to such ATG-1639 polynucleotides.

ATG-1639 of the invention is structurally related to other proteins of the secreted protein ligand for 7-TM receptor, frizzled family, as shown by the results of sequencing the cDNA encoding human ATG-1639. The cDNA sequence of SEQ ID NO:1 contains an open reading frame (nucleotide numbers 200 to 1306) encoding a polypeptide of 369 amino acids of SEQ ID NO:2. The amino acid sequence of Table 1 (SEQ ID NO:2) has about 50.9% identity (using FASTA-P) in 162 amino acid residues with *Xenopus laevis* Frizb [Wang,S., Krinks,M., Lin,K., Luyten,F.P. and Moos,M. Jr. Cell 88 (6), 757-766 (1997)]. The amino acid sequence of Table 1 (SEQ ID NO:2) also has about 49.8% identity with *Xenopus laevis* Frezzled [Leyns,L., Bouwmeester,T., Kim,S.H., Piccolo,S. and De Robertis,E.M. Cell 88 (6), 747-756 (1997)]. The nucleotide sequence of Table 1 (SEQ ID NO:1) has about 81.6% identity (using FASTA-N) in 472 nucleotide residues with mouse sFRP-4 polynucleotide [Rattner,A., Hsieh,J.-C., Smallwood,P.M., Gilbert,D.J., Copeland,N.G., Jenkins,N.A. and Nathans,J. Proc. Natl. Acad. Sci. U.S.A. 94, 2859-2863 (1997)]. The nucleotide sequence of Table 1 (SEQ ID NO:1) also has about 97.9% identity in 328 nucleotide residues with human EST #AA486838 (unknown function) [Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T., Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M., Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B., Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and Waterston,R., The WashU-HHMI Mouse EST Project, Unpublished (1996)].

Table 1^a

TGCGGAGT CCGGGA CTGGAG CTGCCGGG OGGGTT CGCGCCCGAAGG CTGAGAG CTGGCG CTGCTCGT
 GCCCTGTGTGCCAGACGG OGGAG CTCCGCGG COGGACCCCGCGGCCCG CTTTGTG CCGACTGGAGTT
 TGGGGGAAGAACT CTCTGCG CCCAGAGGATTT CTTCTT CGGCGAAGGGACaGCGAAAGATGAGGGT
 GGCAGGAAGAgAAGGG CGCTTT CTGT CTGCCgGGGT CgCag CgCGAGAGGG CAGTGCCATGTT CCTCTC
 CATCCTAgTGGCGCTGTGCCTGTGGCTGCACCTGGCGCTGGGCGTGCGCGGCGCGCCCTGCGAGGCGGT
 GCGCATCCCTATGTGCOGGCAGTGCCTGGAACTCATACGCGGATGCCCAACCACTGCACCACAGCAC
 GCAGGAGAAOGCCAT CCTGGCCAT CGAGCAGTACGAGGAGCTGGTGGACGTGAACTGCAGCGCGTGCT
 GCGCTTCTTCTCTGTGCCATGTACGCGCCATTTGCACCCTGGAGTTCTTGACGACCCTATCAAGCC
 GTGCAAGTGGTGTGCCAAGCGCGCGCGAAGACTGCGAGCCCCTCATGAAGATGTACAACCAAGCTG
 GCCCGAAAGCCTGGCCTGCGACGAgCTGCCTGTCTATGACGCTGGCGTGTGCATCTCGCCTGAAGCCAT
 CGTCACGGACCTCCCGGAgGATGTTAAGTGGATAgACATCACACCAgACATGATGGTAAGGAAAGGCC
 TCTTGATGTTGACTGTAAAGCCTAAgCCCGATGGTGCAAGTGTaAAAAGGTGAAGCCAACCTTTGGC
 AAGTATCTCAGCAAAAaCTACAGCTATGTTATTCATGCCAAATAAAAGCTGTGCAGAGGAGTGGCTG
 CAATGAGGTCACAACGGTGGTGGATGTAAAAGAGATCTTCAAGTCTCATCACCCATCCCTOGAACTCA
 AGTCCGCTCATTACAAATTCTTCTGCCAGTGTCCACACATCTGCCCCATCAAGATGTTCTCATCAT
 GTGTTACGAGTGGCGTTCAAGGATGATGCTTCTTGAAAATTGCTTAGTTGAAAAATGGAGAGATCAGCT

Table 2^b

MRVAGREG RFLSAGVAAREG SAMFLSILVALCLWLHLALGV RGAPCEAVRIPMCRHMPWNIT RMPNHLH
 HSTQENAILAIEQYEELVDVNCsAVL RFFLCAMYAPICTLEFLHDP IKPCKSVQQRARDCEPLMKMYN
 HSWPE SLACDELPVYD RGVCI SPEAIVTDLPEDVKWIDITPDMMVQE RPLDVCCKRLSPDRCKCKKVKP
 TLATYLSKNYSYVIHAKIKAVQ RSGCNEVTTVVDVKEIFKSSSPIPRTQVPLITNSSQCPHILPHQDV
 LIMCYEWRSMMLLENCLVEKW RDQLSKRSIQWEE RLQEQRRTVQDKKKTAGRTSRSNPKPKGKPPAP
 KPASPKNIKT RSAQKKT NPKRV

^a An amino acid sequence of a human ATG-1639 (SEQ ID NO: 2).

TAGTAAAAGAT CCATACAGTGGGAAGAGAGGCTGCAGGAACAGCGGAGAACAGTT CAGGACAAGAAGAA
 AACAGCGGGCGCACCAGT OGT AGT AAT CCCCCAAACCAAAGGGAAG CCT CCT GCT CCCAAACCCAG C
 CAGT CCAAGAAGAACATT AAAACT AGGAGT GCCAGAAGAAAA CAAACCGAAAAGAGTGTGAGCTAA
 CTAGTTT CCAAAGCGGAGACTT COGACTT CCTTACAGGATGAGGCTGGGCaTTGCCTGGGACAg CCTaT
 GTTAGGGCATGTGCCCTTG CCCTAACAACT CCCTGCAGTGCTCTTT CATAgAcACATCTTG CAGCATT
 TTT CCTTAA

^a A nucleotide sequence of a human ATG-1639 (SEQ ID NO: 1).

One polynucleotide of the present invention encoding ATG-1639 may be obtained using standard cloning and screening, from a cDNA library derived from mRNA in cells of human adipose tissue and chondrosarcoma, using the expressed sequence tag (EST) analysis (Adams, M.D., *et al. Science* (1991) 252:1651-1656; Adams, M.D. *et al.*, *Nature*, (1992) 355:632-634; Adams, M.D., *et al.*, *Nature* (1995) 377 Supp:3-174). Polynucleotides of the invention can also be obtained from natural sources such as genomic DNA libraries or can be synthesized using well known and commercially available techniques.

The nucleotide sequence encoding ATG-1639 polypeptide of SEQ ID NO:2 may be identical to the polypeptide encoding sequence contained in Table 1 (nucleotide number 200 to 1306 of SEQ ID NO:1), or it may be a sequence, which as a result of the redundancy (degeneracy) of the genetic code, also encodes the polypeptide of SEQ ID NO:2.

When the polynucleotides of the invention are used for the recombinant production of ATG-1639 polypeptide, the polynucleotide may include the coding sequence for the mature polypeptide or a fragment thereof, by itself; the coding sequence for the mature polypeptide or fragment in reading frame with other coding sequences, such as those encoding a leader or secretory sequence, a pre-, or pro- or prepro- protein sequence, or other fusion peptide portions. For example, a marker sequence which facilitates purification of the fused polypeptide can be encoded. In certain preferred embodiments of this aspect of the invention, the marker sequence is a hexahistidine peptide, as provided in the pQE vector (Qiagen, Inc.) and described in Gentz *et al.*, *Proc Natl Acad Sci USA* (1989) 86:821-824, or is an HA tag. The polynucleotide may also contain non-coding 5' and 3' sequences, such as transcribed, non-translated sequences, splicing and polyadenylation signals, ribosome binding sites and sequences that stabilize mRNA.

Further preferred embodiments are polynucleotides encoding ATG-1639 variants comprising the amino acid sequence of ATG-1639 polypeptide of Table 2 (SEQ ID NO:2) in which several, 5-10, 1-5, 1-3, 1-2 or 1 amino acid residues are substituted, deleted or added, in any combination. Among the preferred polynucleotides of the present invention is the polypeptide contained in Table 3 (SEQ ID NO: 3) encoding the amino acid sequence of Table 4 (SEQ ID NO: 4).

Table 3^f

CACAGCAGCAGANAAAGCCAT CCTGGCCAT OGAGCAGTANGAGGAGCTGGTGGACGTGA
 ACTGCAGCGCCNTGCTGCGCTTCTT CCTCTGTGCCATGTACGCGCCCATTTGCACCCCTGG
 AAGTTTCTTG CANGGACCTATCAAAGCGTGCAAGTGGTGTNNCAANGCGCGNGNGAAC
 GNATTGOGAANCCCCCTATGAANGATGTNNAAACCAACANCTGGGCGNANAACTGGCCTT
 GCGANCGAGCTGGCTNGTCTANTACGTTGGCTGTGCANCTCAGCCTGAAGGCATCGTCA
 ANGAACTTCOGGAAGGATGTTAAGTGGGTAGAAATCAACACCANGACATGANGGGTACAA
 GGANAGGCCTCTTGATGTTGGACTGTAAAAGCCTAAGCCCGANTTGGT

^c A partial nucleotide sequence of a human ATG-1639 (SEQ ID NO: 3).

Table 4^d

NAILAIEQXEELVDVNC SAXL RFFLCAMYAPI CTLE

^a A partial amino acid sequence of a human ATG-1639 (SEQ ID NO: 4).

The present invention further relates to polynucleotides that hybridize to the herein above-described sequences. In this regard, the present invention especially relates to polynucleotides which hybridize under stringent conditions to the herein above-described polynucleotides. As herein used, the term "stringent conditions" means hybridization will occur only if there is at least 95% and preferably at least 97% identity between the sequences.

Polynucleotides of the invention, which are identical or sufficiently identical to a nucleotide sequence contained in SEQ ID NO:1 or a fragment thereof (including that of SEQ ID NO:3), may be used as hybridization probes for cDNA and genomic DNA, to isolate full-length cDNAs and genomic clones encoding ATG-1639 polypeptide and to isolate cDNA and genomic clones of other genes that have a high sequence similarity to the ATG-1639 gene. Such hybridization techniques are known to those of skill in the art. Typically these nucleotide sequences are 80% identical, preferably 90% identical, more preferably 95% identical to that of the referent. The probes generally will comprise at least 15 nucleotides. Preferably, such probes will have at least 30 nucleotides and may have at least 50 nucleotides. Particularly preferred probes will range between 30 and 50 nucleotides.

In one embodiment, to obtain a polynucleotide encoding ATG-1639 polypeptide comprises the steps of screening an appropriate library under stringent hybridization conditions with a labeled probe having the SEQ ID NO: 1 or a fragment thereof (including that of SEQ ID NO: 3), and isolating full-length cDNA and genomic clones containing said polynucleotide sequence. Such hybridization techniques are well known to those of skill in the art. Thus in another aspect, ATG-1639 polynucleotides of the present invention further include a nucleotide sequence comprising a nucleotide sequence that hybridize under stringent condition to a nucleotide sequence having SEQ ID NO: 1 or a fragment thereof (including that of SEQ ID NO:3). Also included with ATG-1639 polypeptides are polypeptide comprising amino acid sequence encoded by nucleotide sequence obtained by the above hybridization condition. Stringent hybridization conditions are as defined above or alternatively conditions under overnight incubation at 42°C in a solution comprising: 50% formamide, 5xSSC (150mM NaCl, 15mM trisodium citrate), 50 mM sodium phosphate (pH7.6), 5x Denhardt's solution, 10% dextran sulfate, and 20 microgram/ml denatured, sheared salmon sperm DNA, followed by washing the filters in 0.1x SSC at about 65°C.

The polynucleotides and polypeptides of the present invention may be employed as research reagents and materials for discovery of treatments and diagnostics to animal and human disease.

Vectors, Host Cells, Expression

The present invention also relates to vectors which comprise a polynucleotide or polynucleotides of the present invention, and host cells which are genetically engineered with vectors of the invention and to the production of polypeptides of the invention by recombinant techniques. Cell-free translation systems can also be employed to produce such proteins using RNAs derived from the DNA constructs of the present invention.

For recombinant production, host cells can be genetically engineered to incorporate expression systems or portions thereof for polynucleotides of the present invention. Introduction of polynucleotides into host cells can be effected by methods described in many standard laboratory manuals, such as Davis et al., *BASIC METHODS IN MOLECULAR BIOLOGY* (1986) and Sambrook et al., *MOLECULAR CLONING: A LABORATORY MANUAL*, 2nd Ed., Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y. (1989) such as calcium phosphate transfection, DEAE-dextran mediated transfection, transfection, microinjection, cationic lipid-mediated transfection, electroporation, transduction, scrape loading, ballistic introduction or infection.

Representative examples of appropriate hosts include bacterial cells, such as streptococci, staphylococci, *E. coli*, *Streptomyces* and *Bacillus subtilis* cells; fungal cells, such as yeast cells and *Aspergillus* cells; insect cells such as *Drosophila* S2 and *Spodoptera* Sf9 cells; animal cells such as CHO, COS, HeLa, C127, 3T3, BHK, HEK 293 and Bowes melanoma cells; and plant cells.

A great variety of expression systems can be used. Such systems include, among others, chromosomal, episomal and virus-derived systems, e.g., vectors derived from bacterial plasmids, from bacteriophage, from transposons, from yeast episomes, from insertion elements, from yeast chromosomal elements, from viruses such as baculoviruses, papova viruses, such as SV40, vaccinia viruses, adenoviruses, fowl pox viruses, pseudorabies viruses and retrovi-

ruses, and vectors derived from combinations thereof, such as those derived from plasmid and bacteriophage genetic elements, such as cosmids and phagemids. The expression systems may contain control regions that regulate as well as engender expression. Generally, any system or vector suitable to maintain, propagate or express polynucleotides to produce a polypeptide in a host may be used. The appropriate nucleotide sequence may be inserted into an expression system by any of a variety of well-known and routine techniques, such as, for example, those set forth in Sambrook *et al.*, *MOLECULAR CLONING, A LABORATORY MANUAL* (*supra*).

For secretion of the translated protein into the lumen of the endoplasmic reticulum, into the periplasmic space or into the extracellular environment, appropriate secretion signals may be incorporated into the desired polypeptide. These signals may be endogenous to the polypeptide or they may be heterologous signals.

If the ATG-1639 polypeptide is to be expressed for use in screening assays, generally, it is preferred that the polypeptide be produced at the surface of the cell. In this event, the cells may be harvested prior to use in the screening assay. If ATG-1639 polypeptide is secreted into the medium, the medium can be recovered in order to recover and purify the polypeptide; if produced intracellularly, the cells must first be lysed before the polypeptide is recovered. ATG-1639 polypeptides can be recovered and purified from recombinant cell cultures by well-known methods including ammonium sulfate or ethanol precipitation, acid extraction, anion or cation exchange chromatography, phosphocellulose chromatography, hydrophobic interaction chromatography, affinity chromatography, hydroxylapatite chromatography and lectin chromatography. Most preferably, high performance liquid chromatography is employed for purification. Well known techniques for refolding proteins may be employed to regenerate active conformation when the polypeptide is denatured during isolation and or purification.

Diagnostic Assays

This invention also relates to the use of ATG-1639 polynucleotides for use as diagnostic reagents. Detection of a mutated form of ATG-1639 gene associated with a dysfunction will provide a diagnostic tool that can add to or define a diagnosis of a disease or susceptibility to a disease which results from under-expression, over-expression or altered expression of ATG-1639. Individuals carrying mutations in the ATG-1639 gene may be detected at the DNA level by a variety of techniques.

Nucleic acids for diagnosis may be obtained from a subject's cells, such as from blood, urine, saliva, tissue biopsy or autopsy material. The genomic DNA may be used directly for detection or may be amplified enzymatically by using PCR or other amplification techniques prior to analysis. RNA or cDNA may also be used in similar fashion. Deletions and insertions can be detected by a change in size of the amplified product in comparison to the normal genotype. Point mutations can be identified by hybridizing amplified DNA to labeled ATG-1639 nucleotide sequences. Perfectly matched sequences can be distinguished from mismatched duplexes by RNase digestion or by differences in melting temperatures. DNA sequence differences may also be detected by alterations in electrophoretic mobility of DNA fragments in gels, with or without denaturing agents, or by direct DNA sequencing. See, e.g., Myers *et al.*, *Science* (1985) 230:1242. Sequence changes at specific locations may also be revealed by nuclease protection assays, such as RNase and S1 protection or the chemical cleavage method. See Cotton *et al.*, *Proc Natl Acad Sci USA* (1985) 85: 4397-4401. In another embodiment, an array of oligonucleotides probes comprising ATG-1639 nucleotide sequence or fragments thereof can be constructed to conduct efficient screening of e.g., genetic mutations. Array technology methods are well known and have general applicability and can be used to address a variety of questions in molecular genetics including gene expression, genetic linkage, and genetic variability. (See for example: M.Chee *et al.*, *Science*, Vol 274, pp 610-613 (1996)).

The diagnostic assays offer a process for diagnosing or determining a susceptibility to heart disease, hypertension, cardiovascular diseases, kidney diseases, obesity, insulin resistance, diabetes, CNS diseases, additions. through detection of mutation in the ATG-1639 gene by the methods described.

In addition, heart disease, hypertension, cardiovascular diseases, kidney diseases, obesity, insulin resistance, diabetes, CNS diseases, additions., can be diagnosed by methods comprising determining from a sample derived from a subject an abnormally decreased or increased level of ATG-1639 polypeptide or ATG-1639 mRNA. Decreased or increased expression can be measured at the RNA level using any of the methods well known in the art for the quantitation of polynucleotides, such as, for example, PCR RT-PCR, RNase protection, Northern blotting and other hybridization methods. Assay techniques that can be used to determine levels of a protein, such as an ATG-1639 polypeptide, in a sample derived from a host are well-known to those of skill in the art. Such assay methods include radioimmunoassays, competitive-binding assays, Western Blot analysis and ELISA assays.

Chromosome Assays

The nucleotide sequences of the present invention are also valuable for chromosome identification. The sequence is specifically targeted to and can hybridize with a particular location on an individual human chromosome. The map-

ping of relevant sequences to chromosomes according to the present invention is an important first step in correlating those sequences with gene associated disease. Once a sequence has been mapped to a precise chromosomal location, the physical position of the sequence on the chromosome can be correlated with genetic map data. Such data are found, for example, in V. McKusick, Mendelian Inheritance in Man (available on line through Johns Hopkins University Welch Medical Library). The relationship between genes and diseases that have been mapped to the same chromosomal region are then identified through linkage analysis (coinheritance of physically adjacent genes). The differences in the cDNA or genomic sequence between affected and unaffected individuals can also be determined. If a mutation is observed in some or all of the affected individuals but not in any normal individuals, then the mutation is likely to be the causative agent of the disease.

Antibodies

The polypeptides of the invention or their fragments or analogs thereof, or cells expressing them can also be used as immunogens to produce antibodies immunospecific for the ATG-1639 polypeptides. The term "immunospecific" means that the antibodies have substantially greater affinity for the polypeptides of the invention than their affinity for other related polypeptides in the prior art.

Antibodies generated against the ATG-1639 polypeptides can be obtained by administering the polypeptides or epitope-bearing fragments, analogs or cells to an animal, preferably a nonhuman, using routine protocols. For preparation of monoclonal antibodies, any technique which provides antibodies produced by continuous cell line cultures can be used. Examples include the hybridoma technique (Kohler, G. and Milstein, C., *Nature* (1975) 256:495-497), the trioma technique, the human B-cell hybridoma technique (Kozbor *et al.*, *Immunology Today* (1983) 4:72) and the EBV-hybridoma technique (Cole *et al.*, MONOCLONAL ANTIBODIES AND CANCER THERAPY, pp. 77-96, Alan R. Liss, Inc., 1985).

Techniques for the production of single chain antibodies (U.S. Patent No. 4,946,778) can also be adapted to produce single chain antibodies to polypeptides of this invention. Also, transgenic mice, or other organisms including other mammals, may be used to express humanized antibodies.

The above-described antibodies may be employed to isolate or to identify clones expressing the polypeptide or to purify the polypeptides by affinity chromatography.

Antibodies against ATG-1639 polypeptides may also be employed to treat heart disease, hypertension, cardiovascular diseases, kidney diseases, obesity, insulin resistance, diabetes, CNS diseases, additions., among others.

Vaccines

Another aspect of the invention relates to a method for inducing an immunological response in a mammal which comprises inoculating the mammal with ATG-1639 polypeptide, or a fragment thereof, adequate to produce antibody and/or T cell immune response to protect said animal from heart disease, hypertension, cardiovascular diseases, kidney diseases, obesity, insulin resistance, diabetes, CNS diseases, additions., among others. Yet another aspect of the invention relates to a method of inducing immunological response in a mammal which comprises, delivering ATG-1639 polypeptide via a vector directing expression of ATG-1639 polynucleotide *in vivo* in order to induce such an immunological response to produce antibody to protect said animal from diseases.

Further aspect of the invention relates to an immunological/vaccine formulation (composition) which, when introduced into a mammalian host, induces an immunological response in that mammal to a ATG-1639 polypeptide wherein the composition comprises a ATG-1639 polypeptide or ATG-1639 gene. The vaccine formulation may further comprise a suitable carrier. Since ATG-1639 polypeptide may be broken down in the stomach, it is preferably administered parenterally (including subcutaneous, intramuscular, intravenous, intradermal etc. injection). Formulations suitable for parenteral administration include aqueous and non-aqueous sterile injection solutions which may contain anti-oxidants, buffers, bacteriostats and solutes which render the formulation isotonic with the blood of the recipient; and aqueous and non-aqueous sterile suspensions which may include suspending agents or thickening agents. The formulations may be presented in unit-dose or multi-dose containers, for example, sealed ampoules and vials and may be stored in a freeze-dried condition requiring only the addition of the sterile liquid carrier immediately prior to use. The vaccine formulation may also include adjuvant systems for enhancing the immunogenicity of the formulation, such as oil-in water systems and other systems known in the art. The dosage will depend on the specific activity of the vaccine and can be readily determined by routine experimentation.

Screening Assays

The ATG-1639 polypeptide of the present invention may be employed in a screening process for compounds which activate (agonists) or inhibit activation of (antagonists, or otherwise called inhibitors) the ATG-1639 polypeptide of the

present invention. Thus, polypeptides of the invention may also be used to assess identify agonist or antagonists from, for example, cells, cell-free preparations, chemical libraries, and natural product mixtures. These agonists or antagonists may be natural substrates, ligands, receptors, etc., as the case may be, of the polypeptide of the present invention; or may be structural or functional mimetics of the polypeptide of the present invention. See Coligan *et al.*, *Current Protocols in Immunology* 1(2):Chapter 5 (1991).

ATG-1639 polypeptides are responsible for many biological functions, including many pathologies. Accordingly, it is desirous to find compounds and drugs which stimulate ATG-1639 polypeptide on the one hand and which can inhibit the function of ATG-1639 polypeptide on the other hand. In general, agonists are employed for therapeutic and prophylactic purposes for such conditions as heart disease, hypertension, cardiovascular diseases, kidney diseases, obesity, insulin resistance, diabetes, CNS diseases, additions.. Antagonists may be employed for a variety of therapeutic and prophylactic purposes for such conditions as heart disease, hypertension, cardiovascular diseases, kidney diseases, obesity, insulin resistance, diabetes, CNS diseases, additions..

In general, such screening procedures may involve using appropriate cells which express the ATG-1639 polypeptide or respond to ATG-1639 polypeptide of the present invention. Such cells include cells from mammals, yeast, *Drosophila* or *E. coli*. Cells which express the ATG-1639 polypeptide (or cell membrane containing the expressed polypeptide) or respond to ATG-1639 polypeptide are then contacted with a test compound to observe binding, or stimulation or inhibition of a functional response. The ability of the cells which were contacted with the candidate compounds is compared with the same cells which were not contacted for ATG-1639 activity.

This full length protein can be used for screening any protein receptors that may interact with it. For example, a GST fusion protein can be produced in a bacteria or baculovirus system and used for identifying additional receptors that may bind to ATG-1639 (please see examples). In addition, the full length ATG-1639 protein can be produced in a large amount in bacterial and/or baculovirus, and protein can be iodinated to look for interacting receptors in a classical binding assay. In addition, this protein is also useful in characterizing potential inhibitors that may inhibit the binding of this protein to its 7-TM receptor in a classical competition experiment.

The assays may simply test binding of a candidate compound wherein adherence to the cells bearing the ATG-1639 polypeptide is detected by means of a label directly or indirectly associated with the candidate compound or in an assay involving competition with a labeled competitor. Further, these assays may test whether the candidate compound results in a signal generated by activation of the ATG-1639 polypeptide, using detection systems appropriate to the cells bearing the ATG-1639 polypeptide. Inhibitors of activation are generally assayed in the presence of a known agonist and the effect on activation by the agonist by the presence of the candidate compound is observed.

The ATG-1639 cDNA, protein and antibodies to the protein may also be used to configure assays for detecting the effect of added compounds on the production of ATG-1639 mRNA and protein in cells. For example, an ELISA may be constructed for measuring secreted or cell associated levels of ATG-1639 protein using monoclonal and polyclonal antibodies by standard methods known in the art, and this can be used to discover agents which may inhibit or enhance the production of ATG-1639 (also called antagonist or agonist, respectively) from suitably manipulated cells or tissues.

The ATG-1639 protein may be used to identify membrane bound or soluble receptors, if any, through standard receptor binding techniques known in the art. These include, but are not limited to, ligand binding and crosslinking assays in which the ATG-1639 is labeled with a radioactive isotope (eg 125I), chemically modified (eg biotinylated), or fused to a peptide sequence suitable for detection or purification, and incubated with a source of the putative receptor (cells, cell membranes, cell supernatants, tissue extracts, bodily fluids). Other methods include biophysical techniques such as surface plasmon resonance and spectroscopy. In addition to being used for purification and cloning of the receptor, these binding assays can be used to identify agonists and antagonists of ATG-1639 which compete with the binding of ATG-1639 to its receptors. Standard methods for conducting screening assays are well understood in the art.

Examples of potential ATG-1639 polypeptide antagonists include antibodies or, in some cases, oligonucleotides or proteins which are closely related to the ligands, substrates, receptors, etc., as the case may be, of the ATG-1639 polypeptide, e.g., a fragment of the ligands, substrates, receptors, or small molecules which bind to the polypeptide of the present invention but do not elicit a response, so that the activity of the polypeptide is prevented.

Prophylactic and Therapeutic Methods

This invention provides methods of treating abnormal conditions such as, heart disease, hypertension, cardiovascular diseases, kidney diseases, obesity, insulin resistance, diabetes, CNS diseases, additions., related to both an excess of and insufficient amounts of ATG-1639 polypeptide activity.

If the activity of ATG-1639 polypeptide is in excess, several approaches are available. One approach comprises administering to a subject an inhibitor compound (antagonist) as hereinabove described along with a pharmaceutically acceptable carrier in an amount effective to inhibit the function of the ATG-1639 polypeptide, such as, for example, by blocking the binding of ligands, substrates, etc., or by inhibiting a second signal, and thereby alleviating the abnormal condition. In another approach, soluble forms of ATG-1639 polypeptides still capable of binding the ligand, substrate,

etc. in competition with endogenous ATG-1639 polypeptide may be administered. Typical embodiments of such competitors comprise fragments of the ATG-1639 polypeptide.

In another approach, soluble forms of ATG-1639 polypeptides still capable of binding the ligand in competition with endogenous ATG-1639 polypeptide may be administered. Typical embodiments of such competitors comprise fragments of the ATG-1639 polypeptide.

In still another approach, expression of the gene encoding endogenous ATG-1639 polypeptide can be inhibited using expression blocking techniques. Known such techniques involve the use of antisense sequences, either internally generated or separately administered. See, for example, O'Connor, *J Neurochem* (1991) 56:560 in Oligodeoxynucleotides as Antisense Inhibitors of Gene Expression, CRC Press, Boca Raton, FL (1988). Alternatively, oligonucleotides which form triple helices with the gene can be supplied. See, for example, Lee *et al.*, *Nucleic Acids Res* (1979) 6:3073; Cooney *et al.*, *Science* (1988) 241:456; Dervan *et al.*, *Science* (1991) 251:1360. These oligomers can be administered *per se* or the relevant oligomers can be expressed *in vivo*.

For treating abnormal conditions related to an under-expression of ATG-1639 and its activity, several approaches are also available. One approach comprises administering to a subject a therapeutically effective amount of a compound which activates ATG-1639 polypeptide, i.e., an agonist as described above, in combination with a pharmaceutically acceptable carrier, to thereby alleviate the abnormal condition. Alternatively, gene therapy may be employed to effect the endogenous production of ATG-1639 by the relevant cells in the subject. For example, a polynucleotide of the invention may be engineered for expression in a replication defective retroviral vector, as discussed above. The retroviral expression construct may then be isolated and introduced into a packaging cell transduced with a retroviral plasmid vector containing RNA encoding a polypeptide of the present invention such that the packaging cell now produces infectious viral particles containing the gene of interest. These producer cells may be administered to a subject for engineering cells *in vivo* and expression of the polypeptide *in vivo*. For overview of gene therapy, see Chapter 20, *Gene Therapy and other Molecular Genetic-based Therapeutic Approaches*, (and references cited therein) in Human Molecular Genetics, T Strachan and A P Read, BIOS Scientific Publishers Ltd (1996). Another approach is to administer a therapeutic amount of ATG-1639 polypeptides in combination with a suitable pharmaceutical carrier.

Formulation and Administration

Peptides, such as the soluble form of ATG-1639 polypeptides, and agonists and antagonist peptides or small molecules, may be formulated in combination with a suitable pharmaceutical carrier. Such formulations comprise a therapeutically effective amount of the polypeptide or compound, and a pharmaceutically acceptable carrier or excipient. Such carriers include but are not limited to, saline, buffered saline, dextrose, water, glycerol, ethanol, and combinations thereof. Formulation should suit the mode of administration, and is well within the skill of the art. The invention further relates to pharmaceutical packs and kits comprising one or more containers filled with one or more of the ingredients of the aforementioned compositions of the invention.

Polypeptides and other compounds of the present invention may be employed alone or in conjunction with other compounds, such as therapeutic compounds.

Preferred forms of systemic administration of the pharmaceutical compositions include injection, typically by intravenous injection. Other injection routes, such as subcutaneous, intramuscular, or intraperitoneal, can be used. Alternative means for systemic administration include transmucosal and transdermal administration using penetrants such as bile salts or fusidic acids or other detergents. In addition, if properly formulated in enteric or encapsulated formulations, oral administration may also be possible. Administration of these compounds may also be topical and/or localized, in the form of salves, pastes, gels and the like.

The dosage range required depends on the choice of peptide, the route of administration, the nature of the formulation, the nature of the subject's condition, and the judgment of the attending practitioner. Suitable dosages, however, are in the range of 0.1-100 µg/kg of subject. Wide variations in the needed dosage, however, are to be expected in view of the variety of compounds available and the differing efficiencies of various routes of administration. For example, oral administration would be expected to require higher dosages than administration by intravenous injection. Variations in these dosage levels can be adjusted using standard empirical routines for optimization, as is well understood in the art.

Polypeptides used in treatment can also be generated endogenously in the subject, in treatment modalities often referred to as "gene therapy" as described above. Thus, for example, cells from a subject may be engineered with a polynucleotide, such as a DNA or RNA, to encode a polypeptide *ex vivo*, and for example, by the use of a retroviral plasmid vector. The cells are then introduced into the subject.

Examples

The examples below are carried out using standard techniques, which are well known and routine to those of skill in the art, except where otherwise described in detail. The examples illustrate, but do not limit the invention.

Cloning strategy for ATG-1639:

A partial clone (ATG-1639, Human Genome Sciences (HGS) EST #1742405 corresponding to SEQ ID NO:3) was initially identified through random searches of the Human Genome Sciences data base. This partial clone (410 bp) showed significant homology (69% over 36 aa) to Bovine Frzb. To get the full length cDNA, both 5' and 3' RACE were performed as described (J. Sambrook, E.F. Fritsch and T. Maniatis (1989) A Laboratory Manual Second. Ed. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York). RACE products were cloned into a TA cloning vector (In Vitro-gene Inc.) and bacteria clones were picked and sequenced. A total of 1458 bp were sequenced, this includes an open reading frame encoding a peptide of 369 aa. Fasta analysis shows this peptide to have high homology to Xenopus Frzb.

Example of expression of ATG-1639 in eukaryotic and prokaryotic systems:

Example 1: Expression and Purification of ATG-1639 in *E. coli*

ATG-1639 encodes a secreted human protein. It has a secreted signal peptide at the N-terminal end that will be cleaved. Thus, a prokaryotic expression system (*E. coli*) will be used to generate large quantities of pure ATG-1639 protein with the C-terminal of the bacterial expressed ATG-1639 being tagged with a histidine tag. The detailed experimental procedure is described below.

The bacterial expression vector pQE60 is used for bacterial expression in this example. (QIAGEN, Inc., 9259 Eton Avenue, Chatsworth, CA, 91311). pQE60 encodes ampicillin antibiotic resistance ("Ampr") and contains a bacterial origin of replication ("ori"), an IPTG inducible promoter, a ribosome binding site ("RBS"), six codons encoding histidine residues that allow for affinity purification using nickel-nitrilo-tri-acetic acid ("Ni-NTA") affinity resin sold by QIAGEN, Inc., supra, and suitable single restriction enzyme cleavage sites. These elements are arranged such that an inserted DNA fragment encoding a polypeptide expresses that polypeptide with the six His residues (i.e., a "6 X His tag") covalently linked to the carboxyl terminus of that polypeptide.

The DNA sequence encoding the desired portion of ATG-1639 protein lacking the hydrophobic leader sequence is amplified from the deposited cDNA clone using PCR oligonucleotide primers which anneal to the amino terminal sequences of the desired portion of the ATG-1639 protein and to sequences in the deposited construct 3' to the cDNA coding sequence. Additional nucleotides containing restriction sites to facilitate cloning in the pQE60 vector are added to the 5' and 3' sequences, respectively.

For cloning the mature protein, the 5' primer has the sequence 5' GAATTCATGagggtggcaggaagagaa 3' (SEQ ID NO:5) containing the underlined EcoR I restriction site followed by 21 nucleotides complementary to the amino terminal coding sequence of the mature ATG-1639 sequence in Table 1. One of ordinary skill in the art would appreciate, of course, that the point in the protein coding sequence where the 5' primer begins may be varied to amplify a DNA segment encoding any desired portion of the complete protein shorter or longer than the mature form. The 3' primer has the sequence 5' GAATTCcactcttttcggttgttt 3' (SEQ ID NO: 6) containing the underlined EcoR I restriction site, followed by 21 nucleotides complementary to the 3' end of the coding sequence immediately before the stop codon in the ATG-1639 DNA sequence in Table 1, with the coding sequence aligned with the restriction site so as to maintain its reading frame with that of the six His codons in the pQE60 vector.

The amplified ATG-1639 DNA fragment and the vector pQE60 are digested with EcoR I and the digested DNA fragments are then ligated together. Insertion of the ATG-1639 DNA into the restricted pQE60 vector places the ATG-1639 protein coding region downstream from the IPTG-inducible promoter and in-frame with an initiating AUG and the six histidine codons.

Example 2: Expression and Purification of untagged ATG-1639 in *E. coli*

In case the tag interferes with biological activity, we will also express ATG-1639 without a tag at either end in *E. coli*. The following is the detailed procedure.

The bacterial expression vector pQE60 is used for bacterial expression in this example. (QIAGEN, Inc., 9259 Eton Avenue, Chatsworth, CA, 91311). pQE60 encodes ampicillin antibiotic resistance ("Ampr") and contains a bacterial origin of replication ("ori"), an IPTG inducible promoter, a ribosome binding site ("RBS"), six codons encoding histidine residues that allow affinity purification using nickel-nitrilo-tri-acetic acid ("Ni-NTA") affinity resin sold by QIAGEN, Inc., supra, and suitable single restriction enzyme cleavage sites. These elements are arranged such that a DNA fragment encoding a polypeptide may be inserted in such a way as to produce that polypeptide with the six His residues (i.e., a "6 X His tag") covalently linked to the carboxyl terminus of that polypeptide. However, in this example, the polypeptide coding sequence is inserted such that translation of the six His codons is prevented and, therefore, the polypeptide is produced with no 6 X His tag.

The DNA sequence encoding the desired portion of the ATG-1639 protein lacking the hydrophobic leader sequence is amplified from the deposited cDNA clone using PCR oligonucleotide primers which anneal to the amino terminal sequences of the desired portion of the ATG-1639 protein and to sequences in the deposited construct 3' to the cDNA coding sequence. Additional nucleotides containing restriction sites to facilitate cloning in the pQE60 vector are added to the 5' and 3' sequences, respectively.

For cloning the mature protein, the 5' primer has the sequence 5' GAATTCATGagggtggcaggaagagaa 3' (SEQ ID NO:5) containing the underlined EcoRI restriction site followed by 21 nucleotides complementary to the amino terminal coding sequence of the mature ATG-1639 sequence in Table 1. One of ordinary skill in the art would appreciate, of course, that the point in the protein coding sequence where the 5' primer begins may be varied to amplify a desired portion of the complete protein shorter or longer than the mature form. The 3' primer has the sequence 5' GAATTCCTA-cactcttttcgggtttgt 3' (SEQ ID NO: 7) containing the underlined EcoRI restriction site followed by 21 nucleotides complementary to the 3' end of the coding sequence immediately before the stop codon in the ATG-1639 DNA sequence in Table 1.

The amplified ATG-1639 DNA fragments and the vector pQE60 are digested with EcoRI and the digested DNA fragments are then ligated together. Insertion of the ATG-1639 DNA into the restricted pQE60 vector places the ATG-1639 protein coding region including its associated stop codon downstream from the IPTG-inducible promoter and in-frame with an initiating AUG. The associated stop codon prevents translation of the six histidine codons downstream of the insertion point.

The ligation mixture is transformed into competent *E. coli* cells using standard procedures such as those described in Sambrook et al., *Molecular Cloning: a Laboratory Manual*, 2nd Ed.; Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY (1989). *E. coli* strain M15/rep4, containing multiple copies of the plasmid pREP4, which expresses the lac repressor and confers kanamycin resistance ("Kan^r"), is used in carrying out the illustrative example described herein. This strain, which is only one of many that are suitable for expressing ATG-1639 protein, is available commercially from QIAGEN, Inc., supra. Transformants are identified by their ability to grow on LB plates in the presence of ampicillin and kanamycin. Plasmid DNA is isolated from resistant colonies and the identity of the cloned DNA confirmed by restriction analysis, PCR and DNA sequencing.

Clones containing the desired constructs are grown overnight ("O/N") in liquid culture in LB media supplemented with both ampicillin (100 mg/ml) and kanamycin (25 mg/ml). The O/N culture is used to inoculate a large culture, at a dilution of approximately 1:25 to 1:250. The cells are grown to an optical density at 600 nm ("OD600") of between 0.4 and 0.6. Isopropyl-b-D-thiogalactopyranoside ("IPTG") is then added to a final concentration of 1 mM to induce transcription from the lac repressor sensitive promoter, by inactivating the lacI repressor. Cells subsequently are incubated further for 3 to 4 hours. Cells then are harvested by centrifugation.

The cells are then stirred for 3-4 hours at 4°C in 6M guanidine-HCl, pH8. The cell debris is removed by centrifugation, and the supernatant containing the ATG-1639 is dialyzed against 50 mM Na-acetate buffer pH6, supplemented with 200 mM NaCl. Alternatively, the protein can be successfully refolded by dialyzing it against 500 mM NaCl, 20% glycerol, 25 mM Tris/HCl pH7.4, containing protease inhibitors. After renaturation, the protein can be purified by ion exchange, hydrophobic interaction and size exclusion chromatography. Alternatively, an affinity chromatography step such as an antibody column can be used to obtain pure ATG-1639 protein. The purified protein is stored at 4°C or frozen at -80°C.

Example 3: Cloning and Expression of ATG-1639 protein in a Baculovirus Expression System

In this illustrative example, the plasmid shuttle vector pA2 is used to insert the cloned DNA encoding the complete protein, including its naturally associated secretory signal (leader) sequence, into a baculovirus to express the mature ATG-1639 protein, using standard methods as described in Summers et al., *A Manual of Methods for Baculovirus Vectors and Insect Cell Culture Procedures*, Texas Agricultural Experimental Station Bulletin No. 1555 (1987). This expression vector contains the strong polyhedrin promoter of the *Autographa californica* nuclear polyhedrosis virus (AcMNPV) followed by convenient restriction sites such as BamHI and Asp718. The polyadenylation site of the simian virus 40 ("SV40") is used for efficient polyadenylation. For easy selection of recombinant virus, the plasmid contains the beta-galactosidase gene from *E. coli* under control of a weak *Drosophila* promoter in the same orientation, followed by the polyadenylation signal of the polyhedrin gene. The inserted genes are flanked on both sides by viral sequences for cell-mediated homologous recombination with wild-type viral DNA to generate viable virus that express the cloned polynucleotide.

Many other baculovirus vectors could be used in place of the vector above, such as pAc373, pVL941 and pAcIM1, as one skilled in the art would readily appreciate, as long as the construct provides appropriately located signals for transcription, translation, secretion and the like, including a signal peptide and an in-frame AUG as required. Such vectors are described, for instance, in Luckow et al., *Virology* 170:31-39.

The cDNA sequence encoding the full length ATG-1639 protein in the deposited clone, including the AUG initiation

codon and the naturally associated leader sequence shown in Table 1 (SEQ ID NO:2), is amplified using PCR oligonucleotide primers corresponding to the 5' and 3' sequences of the gene. The 5' primer has the sequence 5' GAATTCAA-GATGagggtggcaggaagagaa 3' (SEQ ID NO:8) containing the underlined EcoRI restriction enzyme site, an efficient signal for initiation of translation in eukaryotic cells, as described by Kozak, M., J. Mol. Biol. 196:947-950 (1987), followed by 21 bases of the sequence of the complete ATG-1639 protein shown in Table 1, beginning with the AUG initiation codon. The 3' primer has the sequence 5' GAATTCCTAcactctttcgggtttgt 3' (SEQ ID NO:7) containing the underlined EcoRI restriction site followed by 21 nucleotides complementary to the 3' end of the coding sequence immediately before the stop codon in the ATG-1639 DNA sequence in Table 1.

The amplified fragment is isolated from a 1% agarose gel using a commercially available kit ("Geneclean," BIO 101 Inc., La Jolla, Ca.). The fragment then is digested with EcoRI and again is purified on a 1% agarose gel. This fragment is designated herein "F1."

The plasmid is digested with the restriction enzymes EcoRI and optionally, can be dephosphorylated using calf intestinal phosphatase, using routine procedures known in the art. The DNA is then isolated from a 1% agarose gel using a commercially available kit ("Geneclean" BIO 101 Inc., La Jolla, Ca.). This vector DNA is designated herein "V1".

Fragment F1 and the dephosphorylated plasmid V1 are ligated together with T4 DNA ligase. E. coli HB101 or other suitable E. coli hosts such as XL-1 Blue (Stratagene Cloning Systems, La Jolla, CA) cells are transformed with the ligation mixture and spread on culture plates. Bacteria are identified that contain the plasmid with the human ATG-1639 gene using the PCR method, in which one of the primers that is used to amplify the gene and the second primer is from well within the vector so that only those bacterial colonies containing the ATG-1639 gene fragment will show amplification of the DNA. The sequence of the cloned fragment is confirmed by DNA sequencing. This plasmid is designated herein pBacATG-1639.

Five mg of the plasmid pBacATG-1639 is co-transfected with 1.0 mg of a commercially available linearized baculovirus DNA ("BaculoGold[®] baculovirus DNA", Pharmingen, San Diego, CA.), using the lipofection method described by Felgner et al., Proc. Natl. Acad. Sci. USA 84:7413-7417 (1987). 1 mg of BaculoGold[®] virus DNA and 5 mg of the plasmid pBacATG-1639 are mixed in a sterile well of a microtiter plate containing 50 µl of serum-free Grace's medium (Life Technologies Inc., Gaithersburg, MD). Afterwards, 10 µl Lipofectin plus 90 µl Grace's medium are added, mixed and incubated for 15 minutes at room temperature. Then, the transfection mixture is added drop-wise to Sf9 insect cells (ATCC CRL 1711) seeded in a 35 mm tissue culture plate with 1 ml Grace's medium without serum. The plate is rocked back and forth to mix the newly added solution. The plate is then incubated for 5 hours at 27°C. After 5 hours, the transfection solution is removed from the plate and 1 ml of Grace's insect medium supplemented with 10% fetal calf serum is added. The plate is put back into an incubator and cultivation is continued at 27°C for four days.

After four days the supernatant is collected and a plaque assay is performed, as described by Summers and Smith, supra. An agarose gel with "Blue Gal" (Life Technologies Inc., Gaithersburg) is used to allow easy identification and isolation of gal-expressing clones, which produce blue-stained plaques. (A detailed description of a "plaque assay" of this type can also be found in the user's guide for insect cell culture and baculovirology distributed by Life Technologies Inc., Gaithersburg, page 9-10). After appropriate incubation, blue stained plaques are picked with the tip of a micropipettor (e.g., Eppendorf). The agar containing the recombinant viruses is then resuspended in a microcentrifuge tube containing 200 µl of Grace's medium and the suspension containing the recombinant baculovirus is used to infect Sf9 cells seeded in 35 mm dishes. Four days later the supernatants of these culture dishes are harvested and then they are stored at 4°C. The recombinant virus is called V-ATG-1639.

To verify the expression of the ATG-1639 gene, Sf9 cells are grown in Grace's medium supplemented with 10% heat inactivated FBS. The cells are infected with the recombinant baculovirus V-ATG-1639 at a multiplicity of infection ("MOI") of about 2. Six hours later the medium is removed and is replaced with SF900 II medium minus methionine and cysteine (available from Life Technologies Inc., Rockville, MD). If radiolabeled proteins are desired, 42 hours later, 5 mCi of 35S-methionine and 5 mCi 35S-cysteine (available from Amersham) are added. The cells are further incubated for 16 hours and then they are harvested by centrifugation. The proteins in the supernatant as well as the intracellular proteins are analyzed by SDS-PAGE followed by autoradiography (if radiolabeled). Microsequencing of the amino acid sequence of the amino terminus of purified protein may be used to determine the amino terminal sequence of the mature protein and thus the cleavage point and length of the secretory signal peptide.

Example 4: Cloning and Expression of ATG-1639 in Mammalian Cells

A typical mammalian expression vector contains the promoter element, which mediates the initiation of transcription of mRNA, the protein coding sequence, and signals required for the termination of transcription and polyadenylation of the transcript. Additional elements include enhancers, Kozak sequences and intervening sequences flanked by donor and acceptor sites for RNA splicing. Highly efficient transcription can be achieved with the early and late promoters from SV40, the long terminal repeats (LTRS) from Retroviruses, e.g., RSV, HTLV, HIV and the early promoter of the cytomegalovirus (CMV). However, cellular elements can also be used (e.g., the human actin promoter). Suitable

expression vectors for use in practicing the present invention include, for example, vectors such as PSVL and PMSG (Pharmacia, Uppsala, Sweden), pRSVcat (ATCC 37152), pSV2dhfr (ATCC 37146) and pBC12MI (ATCC 67109). Mammalian host cells that could be used include, human Hela 293, H9 and Jurkat cells, mouse NIH3T3 and C127 cells, Cos 1, Cos 7 and CV 1, quail QC1-3 cells, mouse L cells and Chinese hamster ovary (CHO) cells.

Alternatively, the gene can be expressed in stable cell lines that contain the gene integrated into a chromosome. The co-transfection with a selectable marker such as dhfr, gpt, neomycin, or hygromycin allows the identification and isolation of the transfected cells.

The transfected gene can also be amplified to express large amounts of the encoded protein. The DHFR (dihydrofolate reductase) marker is useful to develop cell lines that carry several hundred or even several thousand copies of the gene of interest. Another useful selection marker is the enzyme glutamine synthase (GS) (Murphy et al., *Biochem J.* 227:277-279 (1991); Bebbington et al., *Bio/Technology* 10:169-175 (1992)). Using these markers, the mammalian cells are grown in selective medium and the cells with the highest resistance are selected. These cell lines contain the amplified gene(s) integrated into a chromosome. Chinese hamster ovary (CHO) and NSO cells are often used for the production of proteins.

The expression vectors pC1 and pC4 contain the strong promoter (LTR) of the Rous Sarcoma Virus (Cullen et al., *Molecular and Cellular Biology*, 438447 (March, 1985)) plus a fragment of the CMV-enhancer (Boshart et al., *Cell* 41:521-530 (1985)). Multiple cloning sites, e.g., with the restriction enzyme cleavage sites BamHI, XbaI and Asp718, facilitate the cloning of the gene of interest. The vectors contain in addition the 3' intron, the polyadenylation and termination signal of the rat preproinsulin gene.

Example 4(a): Cloning and Expression in COS Cells

The expression plasmid, pATG-1639 HA, is made by cloning a cDNA encoding ATG-1639 into the expression vector pcDNAI/Amp or pcDNAIII (which can be obtained from Invitrogen, Inc.).

The expression vector pcDNAI/amp contains: (1) an E. coli origin of replication effective for propagation in E. coli and other prokaryotic cells; (2) an ampicillin resistance gene for selection of plasmid-containing prokaryotic cells; (3) an SV40 origin of replication for propagation in eukaryotic cells; (4) a CMV promoter, a polylinker, an SV40 intron; (5) several codons encoding a hemagglutinin fragment (i.e., an "HA" tag to facilitate purification) followed by a termination codon and polyadenylation signal arranged so that a cDNA can be conveniently placed under expression control of the CMV promoter and operably linked to the SV40 intron and the polyadenylation signal by means of restriction sites in the polylinker. The HA tag corresponds to an epitope derived from the influenza hemagglutinin protein described by Wilson et al., *Cell* 37:767 (1984). The fusion of the HA tag to the target protein allows for easy detection and recovery of the recombinant protein with an antibody that recognizes the HA epitope. pcDNAIII contains, in addition, the selectable neomycin marker.

A DNA fragment encoding the ATG-1639 is cloned into the polylinker region of the vector so that recombinant protein expression is directed by the CMV promoter. The plasmid construction strategy is as follows. The ATG-1639 cDNA of the deposited clone is amplified using primers that contain convenient restriction sites, much as described above for construction of vectors for expression of ATG-1639 in E. coli. Suitable primers include the following, which are used in this example. The 5' primer, containing the underlined EcoRI site, a Kozak sequence, an AUG start codon and 6 codons of the 5' coding region of the complete ATG-1639 has the following sequence: 5' GAATTCAGATGagggtggcaggaagagaa 3' (SEQ ID NO: 8). The 3' primer, containing the underlined EcoRI site, a stop codon, and 21 bp of 3' coding sequence has the following sequence (at the 3' end): 5' GAATTCCTAcactctttcgggtttgt 3' (SEQ ID NO: 7).

The PCR amplified DNA fragment and the vector, pcDNAI/Amp, are digested with EcoRI and then ligated. The ligation mixture is transformed into E. coli strain SURE (available from Stratagene Cloning Systems, 11099 North Torrey Pines Road, La Jolla, CA 92037), and the - transformed culture is plated on ampicillin media plates which then are incubated to allow growth of ampicillin resistant colonies. Plasmid DNA is isolated from resistant colonies and examined by restriction analysis or other means for the presence of the ATG1639-encoding fragment.

For expression of recombinant ATG-1639 COS cells are transfected with an expression vector, as described above, using DEAE-DEXTRAN, as described, for instance, in Sambrook et al., *Molecular Cloning: a Laboratory Manual*, Cold Spring Laboratory Press, Cold Spring Harbor, New York (1989). Cells are incubated under conditions for expression of ATG-1639 by the vector.

Expression of the ATG-1639-HA fusion protein is detected by radiolabeling and immunoprecipitation, using methods described in, for example Harlow et al., *Antibodies: A Laboratory Manual*, 2nd Ed.; Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York (1988). To this end, two days after transfection, the cells are labeled by incubation in media containing 35S-cysteine for 8 hours. The cells and the media are collected, and the cells are washed and lysed with detergent-containing RIPA buffer: 150 mM NaCl, 1% NP-40, 0.1% SDS, 0.5% DOC, 50 mM TRIS, pH 7.5, as described by Wilson et al. cited above. Proteins are precipitated from the cell lysate and from the culture media using an HA-specific monoclonal antibody. The precipitated proteins then are analyzed by SDS-PAGE and autoradiography.

An expression product of the expected size is seen in the cell lysate, which is not seen in negative controls.

Example 4(b): Cloning and Expression in CHO Cells

5 The vector pC4 is used for the expression of ATG-1639 protein. Plasmid pC4 is a derivative of the plasmid pSV2-dhfr (ATCC Accession No. 37146). The plasmid contains the mouse DHFR gene under control of the SV40 early promoter. Chinese hamster ovary- or other cells lacking dihydrofolate activity that are transfected with these plasmids can be selected by growing the cells in a selective medium (alpha minus MEM, Life Technologies) supplemented with the chemotherapeutic agent methotrexate. The amplification of the DHFR genes in cells resistant to methotrexate (MTX) 10 has been well documented (see, e.g., Alt, F. W., Kellems, R. M., Bertino, J. R., and Schimke, R. T., 1978, J Biol. Chem. 253:1357-1370, Hamlin, J. L. and Ma, C. 1990, Biochem. et Biophys. Acta, 1097:107-143, Page, M. J. and Sydenham, M.A. 1991, Biotechnology 9:64-68). Cells grown in increasing concentrations of MTX develop resistance to the drug by overproducing the target enzyme, DHFR, as a result of amplification of the DHFR gene. If a second gene is linked to the DHFR gene, it is usually co-amplified and over-expressed. It is known in the art that this approach may be used to 15 develop cell lines carrying more than 1,000 copies of the amplified gene(s). Subsequently, when the methotrexate is withdrawn, cell lines are obtained which contain the amplified gene integrated into one or more chromosome(s) of the host cell.

Plasmid pC4 contains for expressing the gene of interest the strong promoter of the long terminal repeat (LTR) of the Rous Sarcoma Virus (Cullen, et al., Molecular and Cellular Biology, March 1985:438-447) plus a fragment isolated 20 from the enhancer of the immediate early gene of human cytomegalovirus (CMV) (Boshart et al., Cell 41:521-530 (1985)). Downstream of the promoter are BamHI, XbaI, and Asp718 restriction enzyme cleavage sites that allow integration of the genes. Behind these cloning sites the plasmid contains the 3' intron and polyadenylation site of the rat preproinsulin gene. Other high efficiency promoters can also be used for the expression, e.g., the human b-actin promoter, the SV40 early or late promoters or the long terminal repeats from other retroviruses, e.g., HIV and HTLV1. Clon- 25 tech's Tet-Off and Tet-On gene expression systems and similar systems can be used to express the ATG-1639 in a regulated way in mammalian cells (Gossen, M., & Bujard, H. 1992, Proc. Natl. Acad. Sci. USA 89: 5547-5551). For the polyadenylation of the mRNA other signals, e.g., from the human growth hormone or globin genes can be used as well. Stable cell lines carrying a gene of interest integrated into the chromosomes can also be selected upon co-transfection with a selectable marker such as gpt, G418 or hygromycin. It is advantageous to use more than one selectable marker 30 in the beginning, e.g., G418 plus methotrexate.

The plasmid pC4 is digested with the restriction enzymes EcoRI and then dephosphorylated using calf intestinal phosphatase by procedures known in the art. The vector is then isolated from a 1% agarose gel.

The DNA sequence encoding the complete ATG-1639 protein including its leader sequence is amplified using PCR oligonucleotide primers corresponding to the 5' and 3' sequences of the gene. The 5' primer has the sequence 5' GAAT- 35 TCAAGATGaggggtggcaggaagagaa 3' (SEQ ID NO: 8) containing the underlined EcoRI restriction enzyme site followed by an efficient signal for initiation of translation in eukaryotes, as described by Kozak, M., J. Mol. Biol. 196:947-950 (1987), and 21 bases of the coding sequence of ATG-1639 shown in Table 1 (SEQ ID NO:1). The 3' primer has the sequence 5' GAATTCCTAcactcttttcgggttgt 3' (SEQ ID NO: 7) containing the underlined EcoRI restriction site followed 21 nucleotides complementary to the 3' region of the ATG-1639 gene shown in Table 1 (SEQ ID NO:1).

40 The amplified fragment is digested with the endonuclease EcoRI and then purified again on a 1% agarose gel. The isolated fragment and the dephosphorylated vector are then ligated with T4 DNA ligase. E. coli HB101 or XL-1 Blue cells are then transformed and bacteria are identified that contain the fragment inserted into plasmid pC4 using, for instance, restriction enzyme analysis.

Chinese hamster ovary cells lacking an active DHFR gene are used for transfection. 5 mg of the expression plas- 45 mid pC4 is cotransfected with 0.5 mg of the plasmid pSV2-neo using lipofectin (Felgner et al., supra). The plasmid pSV2neo contains a dominant selectable marker, the neo gene from Tn5 encoding an enzyme that confers resistance to a group of antibiotics including G418. The cells are seeded in alpha minus MEM supplemented with 1 mg/ml G418. After 2 days, the cells are trypsinized and seeded in hybridoma cloning plates (Greiner, Germany) in alpha minus MEM supplemented with 10, 25, or 50 ng/ml of methotrexate plus 1 mg/ml G418. After about 10-14 days single clones are 50 trypsinized and then seeded in 6-well petri dishes or 10 ml flasks using different concentrations of methotrexate (50 nM, 100 nM, 200 nM, 400 nM, 800 nM). Clones growing at the highest concentrations of methotrexate are then transferred to new 6-well plates containing even higher concentrations of methotrexate (1 mM, 2 mM, 5 mM, 10 mM, 20 mM). The same procedure is repeated until clones are obtained which grow at a concentration of 100 - 200 mM. Expression of the desired gene product is analyzed, for instance, by SDS-PAGE and Western blot or by reverse phase HPLC analysis. 55

Example 5: Tissue distribution of ATG-1639 mRNA expression

Northern blot analysis is carried out to examine ATG-1639 gene expression in human tissues, using methods

described by, among others, Sambrook et al., cited above. A cDNA probe containing the entire nucleotide sequence of the ATG-1639 protein (SEQ ID NO: 1) is labeled with ^{32}P using the rediprime $\text{\textcircled{O}}$ DNA labeling system (Amersham Life Science), according to manufacturer's instructions. After labeling, the probe is purified using a CHROMA SPIN- 100 $\text{\textcircled{O}}$ column (Clontech Laboratories, Inc.), according to manufacturer's protocol number PT1200-1. The purified labeled probe is then used to examine various human tissues for ATG-1639 mRNA.

Multiple Tissue Northern (MTN) blots containing various human tissues (H) or human immune system tissues (IM) are obtained from Clontech and are examined with the labeled probe using ExpressHyb $\text{\textcircled{O}}$ hybridization solution (Clontech) according to manufacturer's protocol number PT1190-1. Following hybridization and washing, the blots are mounted and exposed to film at -70°C overnight, and films developed according to standard procedures.

SEQUENCE LISTING

5

(1) GENERAL INFORMATION

10

(i) APPLICANT: SmithKline Beecham Corporation

(ii) TITLE OF THE INVENTION: A NOVEL HUMAN GENE SIMILAR
TO A SECRETED PROTEIN FRIZB (ATG-1639)

15

(iii) NUMBER OF SEQUENCES: 8

(iv) CORRESPONDENCE ADDRESS:

20

(A) ADDRESSEE: SmithKline Beecham, Corporate Intellectual
Property

(B) STREET: Two New Horizons Court

(C) CITY: Brentford

25

(D) COUNTY: MIDDLESES

(E) COUNTRY: GB

(F) POST CODE: TW8 9EP

30

(v) COMPUTER READABLE FORM:

(A) MEDIUM TYPE: Diskette

(B) COMPUTER: IBM Compatible

(C) OPERATING SYSTEM: DOS

35

(D) SOFTWARE: FastSEQ for Windows Version 2.0

(vi) CURRENT APPLICATION DATA:

(A) APPLICATION NUMBER: TO BE ASSIGNED

40

(B) FILING DATE: 14-JUL-1997

(C) CLASSIFICATION: UNKNOWN

(vii) PRIOR APPLICATION DATA:

45

(A) APPLICATION NUMBER: 60/047,626

(B) FILING DATE: 23-MAY-1997

50

(viii) ATTORNEY/AGENT INFORMATION:

(A) NAME: CONNELL, Anthony Christopher

55

(B) GENERAL AUTHORISATION NUMBER: 5630

(C) REFERENCE/DOCKET NUMBER: GH-70039

(ix) TELECOMMUNICATION INFORMATION:

(A) TELEPHONE: +44 127 964 4395

(B) TELEFAX: +44 181 975 6294

(C) TELEX:

(2) INFORMATION FOR SEQ ID NO:1:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 1458 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:1:

25	TGCGGAGTCC GGGACTGGAG CTGCCC GGCG GGGTTCGCGC CCCGAAGGCT GAGAGCTGGC	60
	GCTGCTCGTG CCCTGTGTGC CAGACGGCGG AGCTCCGCGG CCGGACCCCG CGGCCCGCT	120
	TTGCTGCCGA CTGGAGTTTG GGGGAAGAAA CTCTCCTGCG CCCAGAGGA TTTCTTCCTC	180
	GGCGAAGGGA CAGCGAAAGA TGAGGGTGGC AGGAAGAGAA GGGCGCTTTC TGTCTGCCGG	240
30	GGTCGCAGCG CGAGAGGGCA GTGCCATGTT CCTCTCCATC CTAGTGGCGC TGTGCCTGTG	300
	GCTGCACCTG GCGCTGGGCG TGCGCGGCGC GCCCTGCGAG GCGGTGCGCA TCCCTATGTG	360
	CCGGCACATG CCCTGGAACA TCACGCGGAT GCCCAACCAC CTGCACCACA GCACGCAGGA	420
	GAACGCCATC CTGGCCATCG AGCAGTACGA GGAGCTGGTG GACGTGAACT GCAGCGCCGT	480
35	GCTGCGCTTC TTCCTCTGTG CCATGTACGC GCCCATTTGC ACCCTGGAGT TCCTGCACGA	540
	CCCTATCAAG CCGTGCAAGT CGGTGTGCCA ACGCGCGCGC GACGACTGCG AGCCCCTCAT	600
	GAAGATGTAC AACCACAGCT GGCCCGAAAG CCTGGCCTGC GACGAGCTGC CTGTCTATGA	660
	CCGTGGCGTG TGCATCTCGC CTGAAGCCAT CGTCACGGAC CTCCCGGAGG ATGTTAAGTG	720
	GATAGACATC ACACCAGACA TGATGGTACA GGAAAGGCCT CTTGATGTTG ACTGTAAACG	780
40	CCTAAGCCCC GATCGGTGCA AGTGTAAGAA GGTGAAGCCA ACTTTGGCAA CGTATCTCAG	840
	CAAAACTAC AGCTATGTTA TTCATGCCAA AATAAAGCT GTGCAGAGGA GTGGCTGCAA	900
	TGAGGTCACA ACGGTGGTGG ATGTAAAGA GATCTTCAAG TCCTCATCAC CCATCCCTCG	960
	AACTCAAGTC CCGCTCATTA CAAATTCTTC TTGCCAGTGT CCACACATCC TGCCCCATCA	1020
45	AGATGTTCTC ATCATGTGTT ACGAGTGGCG TTCAAGGATG ATGCTTCTTG AAAATTGCTT	1080
	AGTTGAAAAA TGGAGAGATC AGCTTAGTAA AAGATCCATA CAGTGGGAAG AGAGGCTGCA	1140
	GGAACAGCGG AGAACAGTTC AGGACAAGAA GAAAACAGCC GGGCGCACCA GTCGTAGTAA	1200
	TCCCCC AAA CCAAAGGGAA AGCTCCTGCT TCCCAAACCA GCCAGTCCCA AGAAGAACAT	1260
50	TAAACTAGG AGTGCCGAGA AGAAAACAAA CCCGAAAAGA GTGTGAGCTA ACTAGTTTCC	1320
	AAAGCGGAGA CTTCCGACTT CTTACAGGA TGAGGCTGGG CATTGCCTGG GACAGCCTAT	1380
	GTTAGGGCAT GTGCCCTTG CCCTAACAAAC TCCCTGCAGT GCTCTTTCAT AGACACATCT	1440

TGCAGCATTT TTCCTTAA

1458

(2) INFORMATION FOR SEQ ID NO:2:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 368 amino acids

(B) TYPE: amino acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:2:

Met Arg Val Ala Gly Arg Glu Gly Arg Phe Leu Ser Ala Gly Val Ala
 1 5 10 15
 Ala Arg Glu Gly Ser Ala Met Phe Leu Ser Ile Leu Val Ala Leu Cys
 20 25 30
 Leu Trp Leu His Leu Ala Leu Gly Val Arg Gly Ala Pro Cys Glu Ala
 35 40 45
 Val Arg Ile Pro Met Cys Arg His Met Pro Trp Asn Ile Thr Arg Met
 50 55 60
 Pro Asn His Leu His His Ser Thr Gln Glu Asn Ala Ile Leu Ala Ile
 65 70 75 80
 Glu Gln Tyr Glu Glu Leu Val Asp Val Asn Cys Ser Ala Val Leu Arg
 85 90 95
 Phe Phe Leu Cys Ala Met Tyr Ala Pro Ile Cys Thr Leu Glu Phe Leu
 100 105 110
 His Asp Pro Ile Lys Pro Cys Lys Ser Val Cys Gln Arg Ala Arg Asp
 115 120 125
 Asp Cys Glu Pro Leu Met Lys Met Tyr Asn His Ser Trp Pro Glu Ser
 130 135 140
 Leu Ala Cys Asp Glu Leu Pro Val Tyr Asp Arg Gly Val Cys Ile Ser
 145 150 155 160
 Pro Glu Ala Ile Val Thr Asp Leu Pro Glu Asp Val Lys Trp Ile Asp
 165 170 175
 Ile Thr Pro Asp Met Met Val Gln Glu Arg Pro Leu Asp Val Asp Cys
 180 185 190
 Lys Arg Leu Ser Pro Asp Arg Cys Lys Cys Lys Lys Val Lys Pro Thr
 195 200 205
 Leu Ala Thr Tyr Leu Ser Lys Asn Tyr Ser Tyr Val Ile His Ala Lys
 210 215 220
 Ile Lys Ala Val Gln Arg Ser Gly Cys Asn Glu Val Thr Thr Val Val
 225 230 235 240
 Asp Val Lys Glu Ile Phe Lys Ser Ser Ser Pro Ile Pro Arg Thr Gln

245 250 255
 Val Pro Leu Ile Thr Asn Ser Ser Cys Gln Cys Pro His Ile Leu Pro
 260 265 270
 5 His Gln Asp Val Leu Ile Met Cys Tyr Glu Trp Arg Ser Arg Met Met
 275 280 285
 Leu Leu Glu Asn Cys Leu Val Glu Lys Trp Arg Asp Gln Leu Ser Lys
 290 295 300
 10 Arg Ser Ile Gln Trp Glu Glu Arg Leu Gln Glu Arg Arg Thr Val
 305 310 315 320
 Gln Asp Lys Lys Lys Thr Ala Gly Arg Thr Ser Arg Ser Asn Pro Pro
 325 330 335
 15 Lys Pro Lys Gly Lys Pro Pro Ala Pro Lys Pro Ala Ser Pro Lys Lys
 340 345 350
 Asn Ile Lys Thr Arg Ser Ala Gln Lys Lys Thr Asn Pro Lys Arg Val
 355 360 365

(2) INFORMATION FOR SEQ ID NO:3:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 409 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:3:

CACAGCACGC AGANAACGCC ATCCTGGCCA TCGAGCAGTA NGAGGAGCTG GTGGACGTGA 60
 ACTGCAGCGC CNTGCTGCGC TTCTTCCTCT GTGCCATGTA CGCGCCCATT TGCACCCTGG 120
 AAGTTTCTTG CANGGACCTA TCAAAGCCGT GCAAGTCGGT GTNNCAANGC GCGNGNGAAC 180
 GNATTGCGAA NCCCCTATGA ANGATGTNNA AACCACANCT GGGCCGNANA ACCTGGCCTT 240
 GCGANCGAGC TGGCTNGTCT ANTACCGTTG GCTGTGCANC TCAGCCTGAA GGCATCGTCA 300
 40 ANGAACCTCC GGAAGGATGT TAAGTGGGTA GAAATCAACA CCANGACATG ANGGGTACAA 360
 GGANAGGCCT CTTGATGTTG GACTGTAAAC GCCTAAGCCC CGANTTGGT 409

(2) INFORMATION FOR SEQ ID NO:4:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 36 amino acids
 (B) TYPE: amino acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:7:

5 GAATTCCTAC ACTCTTTTCG GGTTTGT 27

(2) INFORMATION FOR SEQ ID NO:8:

10 (i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 30 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: single

15 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:8:

20 GAATTCAAGA TGAGGGTGGC AGGAAGAGAA 30

25

Claims

1. An isolated polynucleotide comprising a nucleotide sequence that has at least 80% identity to a nucleotide sequence encoding the ATG-1639 polypeptide of SEQ ID NO:2 over its entire length; or a nucleotide sequence complementary to said isolated polynucleotide.
2. The polynucleotide of claim 1 which is DNA or RNA.
3. The polynucleotide of claim 1 wherein said polynucleotide comprises a nucleotide sequence that has at least 80% identical to that of SEQ ID NO:1 over its entire length.
4. The polynucleotide of claim 3 wherein said nucleotide sequence comprises the ATG-1639 polypeptide encoding sequence contained in SEQ ID NO:1.
5. The polynucleotide of claim 3 which is polynucleotide of SEQ ID NO: 1.
6. A DNA or RNA molecule comprising an expression system, wherein said expression system is capable of producing a ATG-1639 polypeptide comprising an amino acid sequence, which has at least 80% identity with the polypeptide of SEQ ID NO:2 when said expression system is present in a compatible host cell.
7. A host cell comprising the expression system of claim 6.
8. A process for producing a ATG-1639 polypeptide comprising culturing a host of claim 7 under conditions sufficient for the production of said polypeptide and recovering the polypeptide from the culture.
9. A process for producing a cell which produces a ATG-1639 polypeptide thereof comprising transforming or transfecting a host cell with the expression system of claim 6 such that the host cell, under appropriate culture conditions, produces a ATG-1639 polypeptide.
10. A ATG-1639 polypeptide comprising an amino acid sequence which is at least 80% identical to the amino acid sequence of SEQ ID NO:2 over its entire length.

11. The polypeptide of claim 10 which comprises the amino acid sequence of SEQ ID NO:2.

12. An antibody immunospecific for the ATG-1639 polypeptide of claim 10.

5 13. A method for the treatment of a subject in need of enhanced activity or expression of ATG-1639 polypeptide of claim 10 comprising:

10 (a) administering to the subject a therapeutically effective amount of an agonist to said polypeptide; and/or
(b) providing to the subject an isolated polynucleotide comprising a nucleotide sequence that has at least 80% identity to a nucleotide sequence encoding the ATG-1639 polypeptide of SEQ ID NO:2 over its entire length; or a nucleotide sequence complementary to said nucleotide sequence in a form so as to effect production of said polypeptide activity *in vivo*.

15 14. A method for the treatment of a subject having need to inhibit activity or expression of ATG-1639 polypeptide of claim 10 comprising:

20 (a) administering to the subject a therapeutically effective amount of an antagonist to said polypeptide; and/or
(b) administering to the subject a nucleic acid molecule that inhibits the expression of the nucleotide sequence encoding said polypeptide; and/or
(c) administering to the subject a therapeutically effective amount of a polypeptide that competes with said polypeptide for its ligand, substrate, or receptor.

25 15. A process for diagnosing a disease or a susceptibility to a disease in a subject related to expression or activity of ATG-1639 polypeptide of claim 10 in a subject comprising:

30 (a) determining the presence or absence of a mutation in the nucleotide sequence encoding said ATG-1639 polypeptide in the genome of said subject; and/or
(b) analyzing for the presence or amount of the ATG-1639 polypeptide expression in a sample derived from said subject.

35 16. A method for identifying compounds which inhibit (antagonize) or agonize the ATG-1639 polypeptide of claim 10 which comprises:

40 (a) contacting a candidate compound with cells which express the ATG-1639 polypeptide (or cell membrane expressing ATG-1639 polypeptide) or respond to ATG-1639 polypeptide; and
(b) observing the binding, or stimulation or inhibition of a functional response; or comparing the ability of the cells (or cell membrane) which were contacted with the candidate compounds with the same cells which were not contacted for ATG-1639 polypeptide activity.

45 17. An agonist identified by the method of claim 16.

50 18. An antagonist identified by the method of claim 16.

55



European Patent
Office

PARTIAL EUROPEAN SEARCH REPORT

Application Number

which under Rule 45 of the European Patent Convention EP 97 30 8556
shall be considered, for the purposes of subsequent
proceedings, as the European search report

DOCUMENTS CONSIDERED TO BE RELEVANT			
Category	Citation of document with indication, where appropriate, of relevant passages	Relevant to claim	CLASSIFICATION OF THE APPLICATION (Int.Cl.6)
D,Y	A.RATTNER ET AL.: "A family of secreted proteins contains homology to the cysteine-rich ligand- binding domain of frizzled receptors" PROC.NATL.ACAD.SCI., vol. 94, April 1997, pages 2859-63, XP002074862 * figure 1 *	1-16	C12N15/12 C07K14/71 C12P21/00
Y	B.HOANG ET AL.: "Primary Structure and Tissue Distribution of FRZB, a Novel Protein Related to Drosophila Frizzled, Suggest a Role in Skeletal Morphogenesis" J.BIOL.CHEM., vol. 271, no. 42, 1996, pages 26131-7, XP002074863 * figure 1 *	1-16	
A	EMBL database entry HSAA14686, Accession number AA194152, Hillier L. et al., The WashU-Merck EST Project, 24-05-1997 XP002074864 * the whole document *	1-16	
			TECHNICAL FIELDS SEARCHED (Int.Cl.6)
			C07K
INCOMPLETE SEARCH <p>The Search Division considers that the present application, or one or more of its claims, does/do not comply with the EPC to such an extent that a meaningful search into the state of the art cannot be carried out, or can only be carried out partially, for these claims.</p> <p>Claims searched completely :</p> <p>Claims searched incompletely :</p> <p>Claims not searched :</p> <p>Reason for the limitation of the search:</p> <p>see sheet C</p>			
Place of search		Date of completion of the search	Examiner
MUNICH		18 August 1998	Deffner, C-A
CATEGORY OF CITED DOCUMENTS <p>X : particularly relevant if taken alone Y : particularly relevant if combined with another document of the same category A : technological background O : non-written disclosure P : intermediate document</p> <p>T : theory or principle underlying the invention E : earlier patent document, but published on, or after the filing date D : document cited in the application L : document cited for other reasons & : member of the same patent family, corresponding document</p>			

EPO FORM 1503 03 82 (P04C07)

European Patent
Office

PARTIAL EUROPEAN SEARCH REPORT

Application Number
EP 97 30 8556

DOCUMENTS CONSIDERED TO BE RELEVANT			CLASSIFICATION OF THE APPLICATION (Int.Cl.6)
Category	Citation of document with indication, where appropriate, of relevant passages	Relevant to claim	
A	EMBL database entry HS1194675, Accession number AA291847, Hillier L. et al., The WashU-Merck EST Project, 19-05-1997 XP002074865 * the whole document * -----	1-16	
			TECHNICAL FIELDS SEARCHED (Int.Cl.6)

EPO FORM 1503 03.92 (P04C10)



European Patent
Office

INCOMPLETE SEARCH
SHEET C

Application Number
EP 97 30 8556

Although claims 17, 18 are directed to a method of treatment of the human/animal body (Article 52(4) EPC), the search has been carried out and based on the alleged effects of the compound/composition.

Claim(s) searched completely:
1-16

Claim(s) not searched:
17,18

Reason for the limitation of the search (non-patentable invention(s)):

Claims 17 and 18 relate to agonists or antagonists of the protein encoded by the present DNA sequence claimed without any structural information to which such an activity can be linked. Therefore the EPO is not in the position to perform a search for subject-matter of claims 17 and 18.